ELSEVIER

Theoretical Computer Science 186 (1997) 83–105

Theoretical Computer Science

Solving systems of explicit language relations

Ernst L. Leiss

Department of Computer Science, University of Houston, Houston, TX 77204-3475, USA

Received February 1995; revised June 1996
Communicated by M. Nivat

Abstract

While systems of language *equations* have been studied in various contexts, the corresponding problems for general relations between languages have not received much attention. This paper examines relations where the operations involved are unrestricted union and concatenation from the left by a constant. In the case of equations, this defines the classical ones provided each variable has exactly one equation. Since these equations express variables, they will be called explicit. For single explicit relations, we solve the problem of whether there exists a solution, study how to find all solutions, and investigate adequate representations for the solutions. Then we focus on systems of several explicit relations; the questions here are significantly more complicated since for a single variable X, there may be three types of relations, namely

$$X = LX \cup M,$$

$$X \supseteq LX \cup M,$$

$$X \subseteq LX \cup M.$$

We concentrate on decoupled systems; these are systems where for each variable, at most one of the three types may occur (although different variables of the system may occur in relations of different types). Nevertheless, a single variable may have several relations of the same type. We give methods to answer the questions whether there exists a solution, whether there is more than one solution, and how to represent these solutions.

1. Introduction

Equations involving languages have been studied for many years. Yet, the analogous notion of inequality has received relatively little attention. In this paper, we develop a theory of (explicit) language relations that is an analogue of the theory of systems

* Tel.: +1 (713) 743 3359; fax: +1 (713) 743 3335; e-mail: coscel@cs.uh.edu.

PII S0304-3975(96)00218-6

of n (classical or explicit) equations in n unknowns $X_1,\dots,X_n$:

$$X_i = S_{i,1} \cdot X_1 \cup \cdots \cup S_{i,n} \cdot X_n \cup S_{i,0}, \quad i = 1,\dots,n, \tag{1.1}$$

with all $S_{i,j}$ being constant languages over a fixed alphabet A (see [2]). For explicit equations, this is the most general form studied in the literature where the operations involved are unrestricted union and concatenation from the left by a constant language. (The restriction of general concatenation to left-concatenation serves to exclude terms such as XaX which would change the problem substantially.) A solution of this system of explicit equations is then a vector of n languages $(L_1,\dots,L_n)$ over A with the property that substituting the language L_i for every occurrence of the variable X_i (for $i = 1,\dots,n$) in all equations will convert (1.1) into a set of n identities between languages. It is well known that such systems of equations (1.1) always have solutions and that all solutions can be written as regular expressions in terms of the constant languages $S_{i,j}$, regardless of whether the $S_{i,j}$ themselves are regular or not. Furthermore, a solution is unique if for all $i = 1,\dots,n$, there does not exist a sequence $(j_1,\dots,j_t)$ of indices such that $t \geqslant 2$, $j_1 = j_t = i$, and $\lambda \in S_{j_s,j_{s+1}}$ for all $s = 1,\dots,t-1$.

For example, the system

$$X = HX \cup KY, \qquad Y = LX \cup MY \cup N$$

in the variables X and Y has a solution

$$(X,Y) = (H^*K(LH^*K \cup M)^*N,\ (LH^*K \cup M)^*N);$$

note that the languages of this solution are all represented by regular expressions over the "alphabet" $\{H,K,L,M,N\}$ and that the solution for any specific choice of these five languages over the alphabet A is obtained by substituting these languages into the two regular expressions. In particular, it should be clear that none of these five languages need be regular, nor recursively enumerable, for that matter. If none of the languages H, M, and KL contains the empty word, the given solution is unique. If one or more of these three languages contain λ, parametrized regular expressions can be given which represent all possible solutions. Throughout this paper, the parameters in these representations will be denoted by T, possibly subscripted. Thus, if $\lambda \in H$ and $\lambda \in L$, this is a parametrized representation of all solutions with T_X and T_Y (free) parameters ranging over all languages over the underlying alphabet:

$$X = H^*[K(LH^*K \cup M)^*(LH^*T_X \cup N \cup T_Y) \cup T_X],$$

$$Y = (LH^*K \cup M)^*(LH^*T_X \cup N \cup T_Y).$$

Note that for these equations it is sufficient to have at most one parameter per variable. For example, even if additionally $\lambda \in KL$, this parametric representation is the most general one.

Since in the classical equations, the operations are unrestricted union and left-concatenation, we assume the same operations for our explicit language relations. The situation is, however, considerably more complicated since there are now three different types of relations that can hold for any variable; these three types are

equality: $X = LX \cup M$,

super-relation: $X \supseteq LX \cup M$,

sub-relation: $X \subseteq LX \cup M$.

Thus, more than one type may hold for the same variable; moreover, several relations of the same type may hold for the same variable. As an example, consider the system of four explicit relations in the variables X, Y, and Z over the alphabet $\{a, b\}$ (we drop the braces whenever possible):

$$X = a^*X \cup bY \cup a^*,$$

$$X \supseteq bX \cup aY \cup abZ,$$

$$Y \supseteq aX \cup b^*Y \cup bZ,$$

$$Y \subseteq aX \cup bY \cup aZ.$$

Note that both X and Y have two relations (of two different types) while Z has none. It can be shown that this system has a solution (which is not unique) given by

$$(X, Y, Z) = ((a \cup b)^*, b^*a(a \cup b)^*, a(a \cup b)^*).$$

That these three languages are indeed a solution can be easily verified by direct substitution.

We will concentrate in this paper on decoupled systems; these are systems of explicit relations where for every variable X, relations of at most one of the three types $(=, \supseteq, \subseteq)$ hold. This restriction is motivated by the following example of a system of two relations in one variable X over the alphabet $\{a, b\}$:

$$X \subseteq bX \cup b,$$

$$X \supseteq aX \cup a.$$

It is easily seen that no solution can exist since by the first relation (sub), all words in any solution would have to start with b while by the second relation (super), there must be at least one word in a solution that starts with a (namely a itself).

Note that decoupled systems do not stipulate that each variable have only one relation; it is only stipulated that if there are two or more relations, they must all be of the same type. Thus, the following is a decoupled system of explicit language relations in the three variables X, Y, and Z, with the constant languages $\{\lambda\}, \{a\}, \{b\}$, and $\{ab\}$:

$$X = aX \cup bY \cup \lambda,$$

$$X = b^*X \cup abZ,$$

$$Y \supseteq aX \cup bY \cup Z,$$

$$Y \supseteq bX \cup aY \cup aZ.$$

Finally, if we stipulate that each variable in a system of explicit relations have precisely one relation, we call this a strictly decoupled system (note that any such system is automatically decoupled).

2. Definitions

Let A denote the underlying *alphabet*, and let $X_1,\ldots,X_n$ denote n *variables*. We define the class C_n of expressions as follows:

(a) If L is a language over A (also called a *constant*), then L is an expression in C_n. Any variable X_i, for $i \in \{1,\ldots,n\}$, is an expression in C_n.

(b) If α, β are expressions in C_n and L is a language over A, then $\alpha \cup \beta$ (union) and $L \cdot \alpha$ (left-concatenation) are also expressions in C_n.

Concatenation is restricted so that its left operand must be a constant; this will be called *left-concatenation*. Constant languages can be arbitrary; in particular, they need not be regular.

Every expression α in C_n can be viewed as a function from languages to languages:

$$\alpha : (A^*)^n \to A^*.$$

In the following, let $\mathbb{R}$ stand for one of the following *relations* between languages:

equality : $=$

super-relation : $\supseteq$

sub-relation : $\subseteq$

Then a general system of *explicit language relations* over the alphabet A in the n variables $(X_1,\ldots,X_n)$ is defined as follows:

$$X_i \ \mathbb{R}_{i,j} \ \alpha_{i,j} \quad \text{for } i = 1,\ldots,n, \ 1 \leqslant j \leqslant \tau_i \text{ where } \tau_i \geqslant 0 \text{ and } \mathbb{R}_{i,j} \in \{=,\supseteq,\subseteq\}, \ \alpha_{i,j} \in C_n.$$

For each i, τ_i denotes the number of relations the variable X_i has; τ_i may be 0.

If for all i, $\tau_i = 1$ and $\mathbb{R}_{i,1}$ is $=$, we have the *classical language equations*. If for all i, $\mathbb{R}_{i,1} = \mathbb{R}_{i,2} = \cdots = \mathbb{R}_{i,\tau_i}$, we have a *decoupled* system of explicit language relations. If for all i, $\tau_i = 1$, we have a *strictly decoupled* system of explicit language relations.

A system of explicit language relations in the n variables $(X_1,\ldots,X_n)$ has a *solution* $(L_1,\ldots,L_n)$ iff substituting L_i for every occurrence of X_i for all $i = 1,\ldots,n$ yields valid relations between languages:

$$L_i \, \mathbb{R}_{i,j} \, \alpha_{i,j}(L_1,\ldots,L_n) \quad \text{for all } i = 1,\ldots,n \text{ and all } j = 1,\ldots,t_i.$$

A system of *implicit language relations* over the alphabet A in the n variables $(X_1,\ldots,X_n)$ is defined as follows:

$$M_j \, \mathbb{R}_j \, \alpha_j \quad \text{for } j = 1,\ldots,m \text{ where } M_j \subseteq A^*, \ \mathbb{R}_j \in \{=,\supseteq,\subseteq\}, \text{ and } \alpha_j \in C_n \text{ for all } j.$$

Note that m and n are not related. If all $\mathbb{R}_j$ are $=$, this is a system of *implicit language equations*. $(L_1,\ldots,L_n)$ is a solution iff

$$M_j \,\mathbb{R}_j\, \alpha_j(L_1,\ldots,L_n) \quad \text{is valid for all } j=1,\ldots,m.$$

A system of *two-sided language relations* over the alphabet A in the n variables $(X_1,\ldots,X_n)$ is defined as follows:

$$\alpha_j \,\mathbb{R}_j\, \beta_j \quad \text{for } j=1,\ldots,m \text{ where } \mathbb{R}_j \in \{=,\supseteq,\subseteq\} \text{ and } \alpha_j,\beta_j \in C_n \text{ for all } j.$$

If all $\mathbb{R}_j$ are $=$, this is a system of *two-sided language equations*. $(L_1,\ldots,L_n)$ is a solution iff

$$\alpha_j(L_1,\ldots,L_n) \,\mathbb{R}_j\, \beta_j(L_1,\ldots,L_n) \quad \text{is valid for all } j=1,\ldots,m.$$

In the following, we will use explicit language relations as well as implicit and two-sided language equations. Implicit language equations have been studied in [1] where a complete solution technique is given. For two-sided language equations, no general solution techniques are known.

3. Properties of single explicit language relations in one variable

We first study single explicit language relations in a single variable X. For the sake of completeness, we state the classical result for equations. The normal form of this equation is

$$X = LX \cup M$$

and if $\lambda \notin L$, L^*M is the unique solution of this equation; if $\lambda \in L$, $L^*(M \cup T)$ gives a parametric representation of all solutions with T ranging over all languages. Note that neither L nor M (nor T, in the second case) need be regular, even though the solution representation is a regular expression.

Super-relation ($\supseteq$): Here the normal form of a single relation in the variable X is

$$X \supseteq LX \cup M, \tag{3.1}$$

with L and M arbitrary languages over the underlying alphabet A. We can formulate:

Proposition 3.1. *Consider a single explicit super-relation* (3.1) *in the single variable X.*

(a) *L^*M is the minimal solution.*

(b) *A^* is the maximal solution.*

(c) *The set of all solutions is given by the expression $L^*(M \cup T)$ with T ranging over all languages in A^*.*

Proof. It is trivial to verify that L^*M and A^* are indeed solutions and that A^* is the maximal solution. We show now that L^*M is the minimal solution. Suppose there exists a solution S that does not contain L^*M; thus, $L^*M - S \neq \emptyset$. Let w be a shortest word in $L^*M - S$. We know that $w \notin S$, and also $w \notin M$ since $S \supseteq M$ by assumption that S is a solution. Let $A^{<|w|}$ denote the set of all words in A^* strictly shorter than w. Because of the minimality of w,

$$S \cap A^{<|w|} = L^*M \cap A^{<|w|}.$$

Then w must be in LS, $w \in LS$. If $\lambda \notin L$, we have

$$w \in LS \quad \text{iff } w \in L(S \cap A^{<|w|}) = L(L^*M \cap A^{<|w|})$$

and since $w \in L^*M$ and $w \notin M$, we conclude that $w \in LL^*M$, thus $w \in L(L^*M \cap A^{<|w|})$, and hence $w \in LS$. However, this contradicts $S \supseteq LS$ since by assumption $w \notin S$.

If $\lambda \in L$, then $LS = (L - \{\lambda\})S \cup S$. By assumption, $w \notin S$, and therefore

$$w \in LS \quad \text{iff } w \in (L - \{\lambda\})S,$$

which means the argument above (stated for the case $\lambda \notin L$) carries over in its entirety.

Now consider (c). We show that $L^*(M \cup T)$ is a solution:

$$\begin{aligned} L^*(M \cup T) \supseteq L[L^*(M \cup T)] \cup M &= LL^*M \cup M \cup LL^*T \\ &= L^*M \cup LL^*T. \end{aligned}$$

Now consider any word w contained in any solution S, $w \in S$. It follows that $L\{w\} \subseteq LS \subseteq S$, and therefore $L^*\{w\}$ must also be contained in S. Since this holds for any w in S, and since L^*M is minimal, any solution is of the form $L^*(M \cup T)$ for some T over A. □

One sees that in contrast to equations, no differentiation between $\lambda \in L$ and $\lambda \notin L$ is required. It follows from the form of the solution expression $L^*(M \cup T)$ that we can always remove the empty word from L in such a relation since $L^* = (L - \{\lambda\})^*$.

Sub-relation ($\subseteq$): Here the normal form of a single explicit relation in the variable X is

$$X \subseteq LX \cup M, \tag{3.2}$$

with L and M arbitrary languages over A. Let $T^{-1}S$ denote the quotient of S with respect to T: $T^{-1}S = \{w \in A^* \mid s = tw \text{ for words } s \in S \text{ and } t \in T\}$. For any language $S \subseteq L^*M$, define its closure $\mathrm{CL}_{L,M}(S)$ as follows:

$$\mathrm{CL}_{L,M}(S) = S \cup \bigcup_{x \in S - M} (L^*)^{-1}\{x\}.$$

Then we can formulate:

Proposition 3.2. *Consider a single explicit sub-relation* (3.2) *in the single variable X.*

(a) $\emptyset$ *is the minimal solution.*

(b) *If* $\lambda \notin L$, *then* L^*M *is the maximal solution; if* $\lambda \in L$, A^* *is the maximal solution.*

(c) *Assume* $\lambda \notin L$. *For any language* $S \subseteq L^*M$, $\mathrm{CL}_{L,M}(S)$ *is a solution. Conversely, if* S *is a solution,* $S = \mathrm{CL}_{L,M}(S)$.

Proof. Parts (a) and (b) are straightforward. We first show that $C := \mathrm{CL}_{L,M}(S)$ is a solution for any $S \subseteq L^*M$. Let $u \in C$, then either $u \in M$ or $u \in C - M$. Since $S \subseteq L^*M$, by definition of $\mathrm{CL}_{L,M}(S)$, $u \in C - M$ can be written as $u = vu'$ with $v \in L$ and $u' \in C$. Thus $u \in LC \cup M$. Now if S is a solution of $X \subseteq LX \cup M$, then $u \in S$ implies $u \in M$, or $u \in LS$ in which case $u = vu'$ and $v \in L$, $u' \in S$, thus $u' \in \mathrm{CL}_{L,M}(\{x\})$. Since this holds for all $x \in S - M$, $S = \mathrm{CL}_{L,M}(S)$. □

Note that this criterion allows for finite solutions; in fact, this is a major difference between super- and sub-relations. While by Proposition 3.1, a solution of a super-relation is almost always infinite (L^*M is the minimal solution and this language is finite iff $M = \emptyset$ or M is finite and $L \in \{\emptyset, \{\lambda\}\}$), all sub-relations admit finite solutions.

We excluded the case $\lambda \in L$ in Part (c), for the following reason: If $\lambda \in L$, then we have

$$X \subseteq LX \cup M = X \cup (L - \{\lambda\})X \cup M$$

and this implies that every language over A is a solution in X. Consequently, sub-relations (3.2) with $\lambda \in L$ are completely uninteresting. In particular, they can always be removed from any system of relations since they do not restrict the solution space at all.

The remainder of the paper is organized as follows. In the next section, we discuss the question of how to solve a system of explicit *equations* where a variable X_i may have more than one defining equation. This problem turns out to be surprisingly difficult. In Section 5, we show how to solve any strictly decoupled system of explicit language relations. In Section 6, we discuss the reduction of arbitrary decoupled to strictly decoupled systems; we also comment on how to solve general systems and give examples.

4. Solving systems with several explicit equations for one variable

We study the situation where the same variable has several explicit equations of the same type. Recall that this is permitted in decoupled systems. Thus, we are dealing with two (or more) equations for X:

$$X = LX \cup M, \qquad X = PX \cup Q,$$

where as always M and Q represent expressions that may involve variables other than X.

One may be tempted to dismiss this problem as irrelevant and even contradictory, since each equation will give rise to its solution and the resulting two solutions most likely will be incompatible. However, this argument requires closer scrutiny.

If neither L nor P contain the empty word λ and M and Q are constants, then the first equation has the unique solution L^*M while the second has the unique solution P^*Q; an internal inconsistency of the system arises if and only if

$$L^*M \neq P^*Q.$$

Much more interesting is the situation where one or both of L and P contain λ. If both L and P contain λ, we get $L^*(M \cup T)$ and $P^*(Q \cup T')$ as parametrized representation of all solutions. Thus, an internal inconsistency (and therefore proof that no solution can exist) arises only if there do not exist languages T and T' such that

$$L^*(M \cup T) = P^*(Q \cup T').$$

However, in this example there are obviously languages T and T' that satisfy this, namely

$$T = T' = A^*.$$

Interestingly, it is the case where only one of L and P contains the empty word that is more complicated and requires a general discussion.

The main difficulty of the problem is related to multiple solutions. We first formulate a result assuming that X is the only variable:

Proposition 4.1. *Let $X = L_iX \cup M_i$ for $i = 1,\dots,n$ be a system of n equations for the single variable X. Assume that the equations are numbered so that $\lambda \in L_j$ for all $j = 1,\dots,m$ with $0 \leqslant m \leqslant n$. Then the system has a solution in X iff*
(a) *$m = n$ or*
(b) *$m < n$ and*
$(L_j)^*M_j = (L_n)^*M_n$ for all $j = m+1,\dots,n-1$ and there exists a solution in the variables T_s, $s = 1,\dots,m$ of the following system of implicit language equations:

$$(L_s)^*T_s \cup (L_s)^*M_s = (L_n)^*M_n \quad \text{for } s = 1,\dots,m.$$

Proof. Let S be a solution; then clearly $S = L_iS \cup M_i$ for $i = 1,\dots,n$. This implies the conditions of the proposition. Conversely, assume that the conditions hold. If (a) holds then $X = A^*$ is a solution. If (b) holds, then $X = (L_n)^*M_n$ is a solution of the system. □

Corollary 4.2. *Let $X = L_iX \cup M_i$ for $i = 1,\dots,n$ be a system of n equations for the single variable X. If it has a solution, then there exists a representation of all solutions of this system.*

Proof. Applying the formulas $X = (L_i)^*M_i$ if $\lambda \notin L_i$ and $X = (L_i)^*(M_i \cup T_i)$ if $\lambda \in L_i$ for each equation, one obtains n expressions that only contain parameters in addition to constants (but not X). These expressions must all be equal, yielding a system of $n-1$ two-sided language equations. Thus, this system of $n-1$ two-sided language equations in the (at most n) parameters has a solution if and only if the original system of n explicit equations in the variable X has a solution. $\square$

Consider the following two explicit equations over the alphabet $\{a, b\}$ for the variable X:

$$X = a^*X \cup (ab)^*, \qquad X = ab^*X \cup \lambda.$$

The first equation has the parametric solution representation $X = a^*((ab)^* \cup T)$ while the second equation has the uniquely determined solution $X = (ab^*)^*$. Thus, there exists a solution of this system if and only if

$$a^*(ab)^* \cup a^*T = (ab^*)^*.$$

This is a simple example of implicit language equations which were studied in [1]. One can verify (by direct substitution) that this particular implicit language equation has the following solution:

$$T = a(a \cup b)^*.$$

In fact, this implicit language equation has infinitely many regular solutions in the parameter T (see [1]).

Were we to change the second equation to

$$X = ab^*X \cup bb$$

the resulting implicit language equation would be

$$a^*(ab)^* \cup a^*T = (ab^*)^*bb$$

for which no solution exists (see [1], or, more directly, consider any word in $a^*(ab)^*$ that contains at least one b).

These observations point out the need for a more formal treatment of systems of explicit equations in which a variable may have more than one equation (one equation per variable is of course the classical case). Note that all that follows is only relevant if there exists at least one variable with more than one (explicit) equation. If for each variable there is at most one equation, the classical theory applies. (The classical theory assumes exactly one equation per variable; if there are fewer equations than variables, the remaining variables are parameters.)

Let Σ_i be the set of all equations for the variable X_i:

$$\Sigma_i :: X_i = \bigcup_{j=1,\ldots,n} L_{i,j;t} X_j \cup M_{i;t} \quad \text{for } 1 \leqslant t \leqslant \tau_i,$$

where τ_i is the number of equations of the variable X_i. Then the system to be solved is

$$(\Sigma_1, \Sigma_2, \ldots, \Sigma_n).$$

The equation t for X_i is *of type M* iff at least one of the following conditions holds:

(1) There exists a sequence $((i_0,t_0),(i_1,t_1),\dots,(i_r,t_r))$ for some $r\geqslant 1$ such that

$$i_0 = i_r = i, \qquad t_0 = t, \qquad 1 \leqslant t_j \leqslant \tau_{i_j} \quad \text{for } j = 1,\dots,r-1$$

and

$$\lambda \in L_{i_{j-1},i_j;t_{j-1}} \quad \text{for all } j = 1,\dots,r.$$

(2) For some variable X_k of type M, there exists a sequence $((i_0,t_0),(i_1,t_1),\dots,(i_r,t_r))$ for some $r\geqslant 1$ such that

$$i_0 = i, \qquad i_r = k, \qquad t_0 = t, \qquad 1 \leqslant t_j \leqslant \tau_{i_j} \quad \text{for } j = 1,\dots,r-1$$

and

$$L_{i_{j-1},i_j;t_{j-1}} \neq \emptyset \quad \text{for all } j = 1,\dots,r.$$

It is not difficult to verify that one can obtain a parametrized solution expression for X_i if and only if X_i is of type M. Loosely speaking, we say X_i has multiple solutions.

Note that this definition of having multiple solutions in a variable is based on syntactic properties of an equation. It should be pointed out that there are equations that meet our definition of multiplicity but that in fact do not have several solutions. An example is given by the equation in X over the alphabet $\{a,b\}$

$$X = (a \cup b \cup \lambda)X \cup \lambda,$$

which has the parametrized representation of all its solutions

$$X = (a \cup b \cup \lambda)^*(\lambda \cup T) = (a \cup b)^* \cup (a \cup b)^* T,$$

but it is quite obvious that this is equal to $(a \cup b)^*$ regardless of the choice of T; thus there is exactly one solution even though X is of type M.

Note that ultimately, if there is to be any equation of type M for a variable, there has to be an equation of type M for at least one variable according to Part (1) of the definition.

We first reduce a general system $(\Sigma_1,\Sigma_2,\dots,\Sigma_n)$ to one where all variables have at least two equations:

Proposition 4.3. *Consider a general system of equations $(\Sigma_1,\Sigma_2,\dots,\Sigma_n)$. Any variable X_s with $\tau_s = 1$ can be eliminated through syntactic substitution such that the resulting system $(\Sigma_1',\Sigma_2',\dots,\Sigma_{s-1}',\Sigma_{s+1}',\dots,\Sigma_n')$ has a solution in the variables $(X_1,\dots,X_{s-1},X_{s+1},\dots,X_n)$ iff the original system $(\Sigma_1,\Sigma_2,\dots,\Sigma_n)$ has a solution in $(X_1,\dots,X_n)$.*

Proof. If X_s has exactly one equation we can solve it using the standard formula and substitute the resulting expression for any occurrence of X_s in the remaining equations. It is trivial to verify that in the resulting system, the variable X_s does not occur any longer, and that it has a solution iff the original system has a solution. □

One must take care about situations where X_s is of type M. More specifically, if the single equation of X_s is of the form $X_s = LX_s \cup M$ with $\lambda \in L$, then a parameter T_{X_s} will occur in the expression. On the other hand, if X_s is of type M because of condition (2), the multiplicity of the solution will arise when the variable X_k that occurs in that condition yields a parametric representation.

Applying Proposition 4.3 repeatedly and treating variables X_r with $\tau_r = 0$ as parameters yields:

Corollary 4.4. *Any given general system of explicit language equations can be equivalently replaced by a system* $(\Sigma_1, \Sigma_2, \ldots, \Sigma_n)$ *where for all* $i = 1, \ldots, n$, $\tau_i \geqslant 2$.

Assume $n = 1$. There are τ_1 equations for X_1, and $\tau_1 \geqslant 2$. We solve each of the equations individually; this yields the following τ_1 equations:

$$X_i = \begin{cases} (L_{1,1;t})^*(M_{1,1;t} \cup T_{1;t}) & \text{if } \lambda \in L_{1,1;t}, \\ (L_{1,1;t})^* M_{1,1;t} & \text{if } \lambda \notin L_{1,1;t}. \end{cases}$$

Evidently, this is precisely the case of Proposition 4.1.

Let $n \geqslant 2$. Select one variable, say X_s; it has τ_s equations, $\tau_s \geqslant 2$. We distinguish two cases:

(A) If there exists one equation for X_s, say equation t, such that $\lambda \notin L_{s,s;t}$, then replace X_s by the expression

$$(L_{s,s;t})^* \left(\bigcup_{\substack{j=1,\ldots,n \\ j \neq s}} L_{s,j;t} X_j \cup L_{0;t} \right)$$

in all equations of the system $(\Sigma_1', \Sigma_2', \ldots, \Sigma_{s-1}', \Sigma_{s+1}', \ldots, \Sigma_n')$ where Σ_i' is Σ_i with all X_s replaced by the above formula. It follows that in the resulting system, only the $n-1$ variables $X_1, \ldots, X_{s-1}, X_{s+1}, \ldots, X_n$ occur. Assuming now inductively that we have a representation $(\gamma_1, \ldots, \gamma_{s-1}, \gamma_{s+1}, \ldots, \gamma_n)$ of all solutions of $(\Sigma_1', \ldots, \Sigma_{s-1}', \Sigma_{s+1}', \ldots, \Sigma_n')$ in the variables $X_1, \ldots, X_{s-1}, X_{s+1}, \ldots, X_n$ where the γ_i are expressions in which only parameters may occur in addition to constants (but no variables X_j), we then substitute these solution representations into the other equations of Σ_s, i.e., all the equations j for X_s with $j = 1, \ldots, t-1, t+1, \ldots, \tau_s$. Since all these must be equal to

$$(L_{s,s;t})^* \left(\bigcup_{\substack{j=1,\ldots,n \\ j \neq s}} L_{s,j;t} \gamma_j \cup L_{0;t} \right),$$

we obtain a system of two-sided language equations with the property that there is a solution of the original system $(\Sigma_1, \Sigma_2, \ldots, \Sigma_n)$ in the variables $(X_1, \ldots, X_n)$ if and only if this system of two-sided language equations has a solution in the parameters that occur in the γ_i.

(B) If for every equation j for X_s, $\lambda \in L_{s,s;j}$, for $j = 1,\dots,\tau_s$, then choose one of the equations, say equation t. This yields the expression

$$(L_{s,s;t})^* \left(\bigcup_{\substack{j=1,\dots,n \\ j\neq s}} L_{s,j;t}X_j \cup L_{0;t} \cup T_{s;t} \right)$$

for X_s, with $T_{s;t}$ a parameter. Then the remainder of the proof follows analogously, using this expression instead of

$$(L_{s,s;t})^* \left(\bigcup_{\substack{j=1,\dots,n \\ j\neq s}} L_{s,j;t}\gamma_j \cup L_{0;t} \right),$$

In both cases, we reduce the original problem equivalently to one of solving a system of two-sided language equations.

Consider the following system of three equations in the two variables X and Y; we assume that none of the four constant languages L, M, N, and P contains the empty word λ:

$$X = LY, \qquad X = MY, \qquad Y = (\lambda \cup N)Y \cup P.$$

If follows therefore that $Y = N^*P \cup N^*T$ for an arbitrary parameter T; from the first two equations, we obtain

$$X = LN^*P \cup LN^*T, \qquad X = MN^*P \cup MN^*T,$$

and this system has a solution in X iff

$$LN^*P \cup LN^*T = MN^*P \cup MN^*T$$

has a solution in T.

Solving such two-sided language equations is in general an open problem. In specific cases, two-sided equations may turn out to be implicit (i.e., one side is a constant language) in which case, [1] provides a general solution method which moreover provides all solutions.

Here are three examples. First, we consider the following system of four explicit equations in the variables X, Y, and Z, over the alphabet $\{a, b\}$:

(i) $X = a^*X \cup b^*Y \cup aZ$,
(ii) $X = ab^*X \cup \lambda$,
(iii) $Y = (bbb)^*Y$,
(iv) $Z = (aa)^*Z$.

From (ii) we get

$$X = a(a \cup b)^* \cup \lambda \quad (= Q);$$

(iii) and (iv) yield

$$Y = (bbb)^* T_1 \quad \text{and} \quad Z = (aa)^* T_2.$$

Using this, we get from (i)

$$\begin{aligned} X &= a^*(b^* Y \cup aZ \cup T_3) \\ &= a^* b^* (bbb)^* T_1 \cup a^* a(aa)^* T_2 \cup a^* T_3 \\ &= a^* b^* T_1 \cup aa^* T_2 \cup a^* T_3. \end{aligned}$$

Now the resulting implicit language equation is

$$a(a \cup b)^* \cup \lambda = a^* b^* T_1 \cup aa^* T_2 \cup a^* T_3$$

and from [1], we know that there is a solution in (T_1, T_2, T_3), for example,

$$(\emptyset, (a \cup b)^*, \lambda).$$

This solution for (T_1, T_2, T_3) of the implicit language equation results in the following solution of the original system (i)–(iv):

$$Y = \emptyset, \qquad Z = (a \cup b)^*, \qquad X = a^* a(a \cup b)^* \cup a^* = a(a \cup b)^* \cup \lambda \quad (= Q),$$

and it is easy to verify that these three languages do indeed constitute a solution of the original system of equations.

Our second example is the following system of three explicit equations in the variables X and Y, over the alphabet $\{a, b\}$:

(i) $X = aX \cup bY \cup \lambda$,
(ii) $X = ab^* X \cup \lambda$,
(iii) $Y = (bbb)^* Y$.

We derive the implicit language equation

$$a(a \cup b)^* \cup \lambda = a^* b(bbb)^* T \cup a^*$$

and conclude that there is no solution of (i)–(iii) in (X, Y) since there is no solution of the implicit equation in T (see either [1], or consider any $w \in T$ in which case $bw \notin a\ (a \cup b)^* \cup \lambda$ gives a contradiction).

The third example is the following system of three explicit equations in two variables

(i) $X = aY$,
(ii) $X = (aa \cup bb)^* X \cup \lambda$,
(iii) $Y = (ab)^* Y$.

From (iii) we get $Y = (ab)^* T_Y$ which when substituted into (i) yields

$$X = a(ab)^* T_Y.$$

(ii) yields

$$X = (aa \cup bb)^* \cup (aa \cup bb)^* T_X$$

and this in turn yields the following two-sided language equation in the variables T_X and T_Y

$$a(ab)^*T_Y = (aa \cup bb)^* \cup (aa \cup bb)^*T_X.$$

In this particular case, we find that no solution exists since on the right of the equation we have the empty word which is never on the left for any choice of T_X.

One must realize that at several points, choices can be made for parameters, but some choices may result in problems in subsequent steps. For example, in Proposition 4.1(a), A^* is always a solution. However, as this is not the only solution, choosing it will yield a system of equations with one variable less, but this new system may not have a solution even though the original system does (although for a different choice for that variable).

Here is an example: Consider the four equations in two variables over the alphabet $A = \{a, b\}$:

$$X = (aa \cup b)X \cup aY \cup a, \qquad X = (aab \cup b^*)X \cup bY \cup a(ab^*a)^*,$$

$$Y = a^*Y, \qquad Y = b^*Y.$$

From the first equation for X we obtain

$$X = (aa \cup b)^*a(\lambda \cup Y)$$

and from the two equations for Y we get

$$Y = a^*T_{Y;1}, \qquad Y = b^*T_{Y;2}.$$

If we choose $T_{Y;1} = T_{Y;2} = A^*$, then

$$Y = A^* \quad \text{and} \quad X = b^*aA^* \quad \text{(from the first equation for } X\text{)}.$$

However, this yields the following implicit equation (from the second equation for X):

$$b^*aA^* = (aab \cup b^*)^*[a(ab^*a)^* \cup bA^* \cup T_{X;2}]$$

and according to [1], this implicit equation does not have a solution. (This can also be seen directly by considering that b is always contained on the right-hand side, but is obviously not in the left-hand side.) Thus, if $Y = A^*$, the given system of four equations in two variables does not have a solution. However, if one chooses

$$T_{Y;1} = T_{Y;2} = \emptyset,$$

there is a solution of the original system of equations. This choice implies $Y = \emptyset$, and thus we obtain $X = (aa \cup b)^*a$ from the first equation and $X = (aab \cup b^*)^*[a(ab^*a)^* \cup T_{X;2}]$ from the second equation for X. Therefore, we must solve the following implicit equation in $T_{X;2}$:

$$(aa \cup b)^*a = (aab \cup b^*)^*[a(ab^*a)^* \cup T_{X;2}],$$

which has a solution, for example for $T_{X;2} = \emptyset$. One can also verify directly, by substitution, that

$$(X, Y) = ((aa \cup b)^* a, \emptyset)$$

is a solution of the given system of four explicit equations in two variables.

5. Solving strictly decoupled systems

We describe how to solve a system of explicit language relations in which each variable has at most one relation. Note that different variables may have relations of different types. For example, if we have three variables X, Y, and Z, and if we use L_i for the requisite constant languages, we may write

$$X \mathbb{R}_1 L_1X \cup L_2Y \cup L_3Z \cup L_4,$$

$$Y \mathbb{R}_2 L_5X \cup L_6Y \cup L_7Z \cup L_8,$$

$$Z \mathbb{R}_3 L_9X \cup L_{10}Y \cup L_{11}Z \cup L_{12},$$

where each of $\mathbb{R}_1$, $\mathbb{R}_2$, and $\mathbb{R}_3$ represents one of $=$, $\supseteq$ and $\subseteq$. Since it is not necessary that each variable have an equation, there are altogether 63 different (non-empty) strictly decoupled systems in three variables (with indeterminate constant languages).

Theorem 5.1. *Any strictly decoupled system of explicit language relations has a solution. If all constant languages in the system are regular, a regular solution can always be effectively determined.*

Proof. Given a strictly decoupled system

$$X_i \mathbb{R}_i L_{i,1}X_1 \cup \cdots \cup L_{i,n}X_n \cup L_{i,0}, \quad i = 1, \ldots, n, \tag{5.1}$$

with $\mathbb{R}_i \in \{=, \supseteq, \subseteq\}$, we replace it with a system of explicit equations

$$X_i = L_{i,1}X_1 \cup \cdots \cup L_{i,n}X_n \cup L_{i,0}, \quad i = 1, \ldots, n, \tag{5.2}$$

and solve this system of equations. By the classical theory, there is at least one solution $(S_1, \ldots, S_n)$ in $(X_1, \ldots, X_n)$ of this system of equations. This solution is regular if all the constant languages $L_{i,j}$ are. It should now be obvious that any solution of (5.2) is also a solution of the given strictly decoupled system of explicit relations (5.1). Consequently, $(S_1, \ldots, S_n)$ is a solution of (5.1) in $(X_1, \ldots, X_n)$ which is regular if all the constant languages $L_{i,j}$ are. □

While the existence of a solution of a strictly decoupled system is thus shown to be almost trivial, the question of a parametrized representation of all solutions is not. It should be noted that the solution given in Theorem 5.1 is neither minimal nor maximal in general: In replacing super-relations ($\supseteq$) by equations ($=$) we are effectively forcing the variable corresponding to the affected relation to be minimized according to Proposition 3.1, while in replacing sub-relations ($\subseteq$) by equations ($=$) we are effectively forcing the variable corresponding to the affected relation to be maximized according to Proposition 3.2.

Corollary 5.2. *Given a strictly decoupled system* (5.1) *of explicit language relations, it is possible to determine its minimal and its maximal solution in* $(X_1,\dots,X_n)$. *If these are identical,* (5.1) *has a unique solution.*

Proof. Let $i_1,\dots,i_a$ be all those indices for which X_{i_s} has a sub-relation, $a\geqslant 0$:

$$X_{i_s} \subseteq L_{i_s,1}X_1 \cup \cdots \cup L_{i_s,n}X_n \cup L_{i_s,0}.$$

Let $i_{a+1},\dots,i_{a+b}$ be all those indices for which X_{i_s} has a super-relation, $b\geqslant 0$:

$$X_{i_s} \supseteq L_{i_s,1}X_1 \cup \cdots \cup L_{i_s,n}X_n \cup L_{i_s,0}.$$

Finally, let $i_{a+b+1},\dots,i_m$ be all those indices for which X_{i_s} has an equation, $m\leqslant n$:

$$X_{i_s} = L_{i_s,1}X_1 \cup \cdots \cup L_{i_s,n}X_n \cup L_{i_s,0}.$$

Here is how the maximal solution of (5.1) is obtained:

Replace all X_{i_s} with $s=1,\dots,a$ by

$$(L_{i_s,i_s})^*[L_{i_s,1}X_1 \cup \cdots \cup L_{i_s,i_s-1}X_{i_s-1} \cup L_{i_s,i_s+1}X_{i_s+1} \cup \cdots \cup L_{i_s,n}X_n \cup L_{i_s,0}]$$

and all X_{i_s} with $s=a+1,\dots,a+b$ by

$$A^*$$

in all the *equations* (and remove all relations that are not equations). This results in a system Σ of explicit language equations in which each variable X_{i_s} with $s=a+b+1,\dots,m$ has at most one equation. Thus, the system Σ has a parametric representation of all solutions in the variables X_{i_s}, $s=a+b+1,\dots,m$. In these parametric solutions, set all parameters to A^*. This is the maximal solution of Σ. It is now trivial to see that the maximal solution of Σ together with the solutions for the variables X_{i_s} with $s=1,\dots,a$ and with $s=a+1,\dots,a+b$ which were previously fixed (and may contain variables X_{i_s} with $s=a+b+1,\dots,m$, which must be replaced by the languages from the solution of Σ, and which were maximal according to Propositions 3.1 and 3.2) is the unique maximal solution of (5.1).

To determine the unique minimal solution we proceed analogously, substituting the minimal solutions according to Propositions 3.1 and 3.2 for the variables with super- and sub-relations and choosing the empty language for any parameter that occurs in a parametric representation.

It is clear that in the case where the minimal and the maximal solutions are identical, the solution is unique. $\square$

All solutions must "lie between" the minimal and the maximal solutions. For super-relations, Proposition 3.1 indicates that we can choose an arbitrary language T and obtain a new solution (by adding to the minimal solution L^*M the term L^*T). For sub-relations, this connection is not as clear since according to Proposition 3.2, the required closure is more complicated. It is an open problem how to combine the two closure operations of Propositions 3.1 and 3.2 to obtain a representation of all solutions of a given strictly decoupled system of equations.

Consider the following strictly decoupled system of three explicit equations in the three variables X, Y, and Z, over the alphabet $\{a, b, c\}$:

$$X = a^*X \cup b^*Y \cup c^*Z,$$

$$Y \subseteq aaZ,$$

$$Z \supseteq c^*X \cup a^*b^*Z \cup \lambda.$$

According to the proof of Theorem 5.1, we obtain the following system of equations:

$$X = a^*X \cup b^*Y \cup c^*Z,$$

$$Y = aaZ,$$

$$Z = c^*X \cup a^*b^*Z \cup \lambda,$$

which has the following solution:

$$X = a^*(b^*aa \cup c^*)(a \cup b \cup c)^*[c^*a^*T_X \cup \lambda \cup T_Z] \cup a^*T_X,$$

$$Y = aa(a \cup b \cup c)^*[c^*a^*T_X \cup \lambda \cup T_Z],$$

$$Z = (a \cup b \cup c)^*[c^*a^*T_X \cup \lambda \cup T_Z],$$

where the T_X and T_Y are free parameters ranging over all languages.

Here are the maximal and the minimal solutions according to Corollary 5.2:

Maximal: We start by setting $Z = A^*$; this implies $Y = aaA^*$ and thus X is the maximal solution of the equation

$$X = a^*X \cup b^*aaA^* \cup c^*A^*,$$

which is A^*. Thus, we obtain the following maximal solution:

$$(X, Y, Z) = (A^*, aaA^*, A^*).$$

Minimal: We start by setting $Z = (a \cup b)^*(c^*X \cup \lambda)$. It follows that the resulting minimal solution is

$$(X, Y, Z) = (A^*, \emptyset, A^*).$$

6. Solving general decoupled systems

While we have seen that strictly decoupled systems are very easily solved, what differentiates strictly decoupled from general decoupled systems is the possibility of having several relations (of the same type) for a single variable, and this has already proven (in Section 3) to be a major problem for the best understood of our language relations, namely equations. It turns out that for explicit relations, having several defining relations for the same variable is easier to handle, especially for super-relations. This is due to the following

Proposition 6.1. *Assume the following two super-relations hold for the variable X_i:*

$$X_i \supseteq H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0,$$

$$X_i \supseteq K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0.$$

Then these two relations have a solution in $X_1,\ldots,X_n$ if and only if the following relation has a solution in $X_1,\ldots,X_n$:

$$X_i \supseteq (H_1 \cup K_1) \cdot X_1 \cup \cdots \cup (H_n \cup K_n) \cdot X_n \cup (H_0 \cup K_0).$$

Proof. First assume that the two relations are solved by $X_1,\ldots,X_n$. It follows that

$$X_i \cup X_i \supseteq [H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0] \cup [K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0]$$

and this implies

$$X_i \supseteq (H_1 \cup K_1) \cdot X_1 \cup \cdots \cup (H_n \cup K_n) \cdot X_n \cup (H_0 \cup K_0).$$

For the converse, we observe that

$$X_i \supseteq (H_1 \cup K_1) \cdot X_1 \cup \cdots \cup (H_n \cup K_n) \cdot X_n \cup (H_0 \cup K_0)$$
$$= [H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0] \cup [K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0]$$

and from this it follows trivially that

$$X_i \supseteq H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0 \quad \text{and} \quad X_i \supseteq K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0. \qquad \square$$

The significance of this proposition comes from the fact that it allows us to replace several explicit super-relations for the same variable by a single one in such a way that any solution of the original system is also a solution of the resulting system and vice versa. In other words, after applying Proposition 6.1 repeatedly to an arbitrary decoupled system we obtain a decoupled system in which no variable may have more than one super-relation.

We turn now to sub-relations, more specifically to variables with more than one explicit sub-relation. There is no analogue to Proposition 6.1 for these relations.

However, we can state the following result:

Proposition 6.2. *Assume the following two sub-relations hold for the variable* X_i:

$$X_i \subseteq H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0,$$

$$X_i \subseteq K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0.$$

Then these two relations have a solution in $X_1, \ldots, X_n$ *if and only if the following relation has a solution in* $X_1, \ldots, X_n$:

$$X_i \subseteq [H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0] \cap [K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0].$$

Proof. First assume that the two relations are solved by $X_1, \ldots, X_n$. It follows that

$$X_i \cap X_i \subseteq [H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0] \cap [K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0].$$

Conversely,

$$X_i \subseteq [H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0] \cap [K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0]$$

implies

$$X_i \subseteq H_1 \cdot X_1 \cup \cdots \cup H_n \cdot X_n \cup H_0 \quad \text{and} \quad X_i \subseteq K_1 \cdot X_1 \cup \cdots \cup K_n \cdot X_n \cup K_0. \qquad \square$$

This suggests that consolidating several explicit inclusions for the same variable into a single inclusion is unlikely to result in a single inclusion using only union and left-concatenation (note that the representation in Proposition 6.2 requires intersection as well).

Theorem 6.3. *Consider a decoupled system of explicit relations in which no variable may have more than one relation of type =. Then the system always has a solution. Furthermore, if all the constant languages involved in the relations are regular, the system always has a regular solution.*

Proof. By Proposition 3.2, every sub-relation has a universal solution, namely the empty language $\emptyset$. As a result, we can set any variable X defined by one or more inclusions to $\emptyset$, replacing every occurrence of X in the relations by $\emptyset$; by Proposition 6.1, we may assume that a variable has at most one super-relation. The resulting new system has the following property: Any solution of the new system is also a solution of the original system (with the variables defined by inclusions being $\emptyset$). Furthermore by Theorem 5.1, the new system always has a solution, since it is strictly decoupled. The statement about the regularity of a solution is obvious. $\square$

We excluded decoupled systems in which some variable has more than one equation, because we do not know in general how to handle this situation. This was at length discussed in Section 3. However, it should be clear that in many instances, Section 3 does allow us to deal with variables with several defining equations, and in those cases, we can solve the general decoupled system.

The remainder of this section will deal with general systems of explicit language relations. They differ from decoupled systems in that a particular variable may have defining relations of different types. It does not appear that there is a consolidation theorem that permits one to replace several relations for that variable by a single one. While this was possible for super-relations, no such results exist for sub-relations nor for equations. Nevertheless, the techniques outlined in this paper can be used to simplify a given general system and then attempt to solve that simpler system.

Consider the system of two explicit relations in one variable:

$$X \subseteq LX \cup M, \qquad X \supseteq PX \cup Q.$$

Recall that for $\lambda \notin L$, L^*M is the maximal solution for the inclusion, while P^*Q is the minimal solution for the containment, if we consider the relations separately. We have:

Proposition 6.4. *Assume $\lambda \notin L$. If P^*Q is not contained in L^*M, no solution of the system exists.*

Proof. If P^*Q is not contained in L^*M, then consider a word $w \in P^*Q - L^*M$. By definition of P^*Q, w must be contained in any solution of the super-relation $X \supseteq PX \cup Q$. Since L^*M is the maximal solution of the sub-relation $X \subseteq LX \cup M$ and $w \notin L^*M$, no solution of the system can exist. □

It is an open question whether the condition $P^*Q \subseteq L^*M$ is sufficient for a solution to exist.

Consider the system in X over $\{a, b\}$:

$$X \subseteq ab^*X \cup \lambda, \qquad X \supseteq aab^*X \cup a.$$

One can verify that the maximal solution of the first relation is $\lambda \cup a(a \cup b)^*$ while the second relation has the minimal solution $a(ab^*a)^*$; clearly $a(ab^*a)^* \subseteq \lambda \cup a(a \cup b)^*$. One can then verify that the maximal solution of the first relation is also a solution of the second relation. Whether this is always so is an open question.

This suggests that in general there is again no consolidation process available. However, while in the case of decoupled systems we could set all variables defined by sub-relations to the empty language, this is no longer possible. One approach that can be tried is to view a general system as a decoupled system together with identifications of variables. This will lead to systems of implicit equations or to two-sided equations provided we have parametric representations of all solutions.

We conclude this section with two examples. The first example is a system of six explicit relations in three variables over the alphabet $\{a\}$:

(i) $X = (a^5)^*X \cup a^7Y \cup a^3Z \cup \lambda$,
(ii) $X \supseteq a^7X \cup a^5Z$,

(iii) $Y \subseteq aX \cup a^2Y \cup a^4Z$,
(iv) $Z \supseteq a^3X \cup a^5Y \cup a^7Z$,
(v) $Z \supseteq a^5X \cup a^7Y \cup a^3Z$,
(vi) $Z = a^7X \cup a^3Y \cup a^5Z \cup \lambda$.

Consolidating (iv) and (v) yields

$$Z \supseteq (a^3 \cup a^5)X \cup (a^5 \cup a^7)Y \cup (a^3 \cup a^7)Z.$$

Relation (vi) has the following solution expression

$$Z = (a^5)^* a^7 X \cup (a^5)^* a^3 Y \cup (a^5)^*.$$

Relation (i) has a parametrized representation of all its solutions

$$X = (a^5)^* a^7 Y \cup (a^5)^* a^3 Z \cup (a^5)^* T_{X,1} \cup (a^5)^*.$$

Super-relation (ii) has a parametrized representation of all its solutions (by Proposition 3.1):

$$X = (a^7)^* a^5 Z \cup (a^7)^* T_{X,2}.$$

Thus, we now have

$$X = (a^5)^* a^7 Y \cup (a^5)^* a^3 Z \cup (a^5)^* T_{X,1} \cup (a^5)^*,$$
$$X = (a^7)^* a^5 Z \cup (a^7)^* T_{X'},$$
$$Y \subseteq aX \cup a^2Y \cup a^4Z,$$
$$Z = (a^3 \cup a^7)^*(a^3 \cup a^5)X \cup (a^3 \cup a^7)^*(a^5 \cup a^7)Y \cup (a^3 \cup a^7)^* T_Z \quad \text{(by Proposition 3.1)},$$
$$Z = (a^5)^* a^7 X \cup (a^5)^* a^3 Y \cup (a^5)^*.$$

By the last two equations, for a solution to exist, the following equation holds for some choice of T_Z:

$$\begin{aligned}&(a^5)^* a^7 X \cup (a^5)^* a^3 Y \cup (a^5)^* \\ &\quad = (a^3 \cup a^7)^*(a^3 \cup a^5)X \cup (a^3 \cup a^7)^*(a^5 \cup a^7)Y \cup (a^3 \cup a^7)^* T_Z.\end{aligned}$$

Now it follows that the left-hand side contains λ (since $\lambda \in (a^5)^*$) and cannot contain *aaa* (since $aaa \notin (a^5)^* a^7 X$ for any language X, $aaa \notin (a^5)^*$, and $aaa \notin (a^5)^* a^3 Y$ because $\lambda \notin Y$). This, however, leads to a contradiction on the right-hand side: For λ to be contained there, we must have $\lambda \in T_Z$, but this implies that *aaa* must then also be contained in the right-hand side, in contradiction to the observation that *aaa* is not contained in the left-hand side. Therefore, the original system of language relations does not have a solution.

Our second example contains three explicit relations in two unknowns over the alphabet $\{a, b\}$:

(i) $X \subseteq aX \cup bY \cup \lambda$,

(ii) $X \supseteq a^2X \cup b^3Y$,
(iii) $Y = a^*X \cup b^*Y$.
From (iii), we get a parametrized representation of all solutions for Y:

$$Y = b^*a^*X \cup b^*T_Y.$$

Substituting this into (i) and (ii) gives

$$X \subseteq aX \cup bb^*a^*X \cup bb^*T_Y \cup \lambda,$$
$$X \supseteq a^2X \cup b^3b^*a^*X \cup b^3b^*T_Y,$$

and from the super-relation, we get the following parametrized representation of all solutions for X:

$$X = [a^2 \cup b^3b^*a^*]^*b^3b^*T_Y \cup [a^2 \cup b^3b^*a^*]^*T_X.$$

Substituting this into the equation for Y yields

$$Y = b^*a^*[a^2 \cup b^3b^*a^*]^*b^3b^*T_Y \cup b^*a^*[a^2 \cup b^3b^*a^*]^*T_X \cup b^*T_Y.$$

We can now rewrite the sub-relation (for X) as follows:

$$\begin{aligned}&[a^2 \cup b^3b^*a^*]^*b^3b^*T_Y \cup [a^2 \cup b^3b^*a^*]^*T_X\\ &\quad\subseteq (a \cup bb^*a^*)[a^2 \cup b^3b^*a^*]^*b^3b^*T_Y \cup (a \cup bb^*a^*)[a^2 \cup b^3b^*a^*]^*T_X\\ &\qquad \cup bb^*T_Y \cup \lambda.\end{aligned}$$

It is quite obvious that

$$T_X = T_Y = \emptyset$$

is a solution of this relation in T_X and T_Y; this implies the following solution of the original system:

$$X = Y = \emptyset.$$

There are other solutions. For example the choice

$$T_X = T_Y = (a \cup b)^*$$

also yields a solution, since on the left-hand side we have $[a^2 \cup b^3b^*a^*]^*T_X$ which is equal to $(a \cup b)^*$; on the right-hand side we have $bb^*T_Y = b(a \cup b)^*, (a \cup bb^*a^*)[a^2 \cup b^3b^*a^*]^*T_X$ contains $a(a \cup b)^*$, and $(a \cup b)^* = a(a \cup b)^* \cup b(a \cup b)^* \cup \lambda$. This choice for T_X and T_Y yields the following solution of the original system:

$$X = Y = (a \cup b)^*.$$

Thus, these are the minimal and the maximal solutions of the original system of three relations in two variables.

6. Conclusion

We have introduced explicit language relations and derived methods for solving them under certain conditions. However, the general problem remains open. This is mainly due to an apparently innocent generalization of systems of explicit language equations (something assumed to be classical) whereby we permitted several defining equations for the same variable. Far from being classical, this generalization renders systems of equations not to have solutions in some cases and in other cases their solution is equivalent to the solution of two-sided language equations; it is the solution of those equations that is open. Intriguingly, there are major differences between super- and sub-relations. For example, one permits consolidation of several defining explicit relations for the same variable into one, the other does not; one has a parametrized representation of all solutions, the other does not. In many cases, existence of a solution of such a system of relations is reduced to the question whether a system of implicit language equations can be solved; implicit equation have been studied in [1].

References

[1] E.L. Leiss, Implicit language equations: existence and uniqueness of solutions, *Theoret. Comput. Sci. A* **145** (1995) 71–93.

[2] A. Salomaa, *Theory of Automata* (Pergamon Press, Oxford, 1969).